Microscopic Shape from Focus Using Active Illumination *

Minori Noguchi and Shree K. Nayar
Department of Computer Science
Columbia University, New York, NY 10027

Abstract

Shape from focus relies on surface texture for the computation of depth. In many real-world applications, surfaces can be smoothly shaded and lacking in detectable texture. In such cases, shape from focus generates inaccurate and sparse depth maps. This paper presents a novel extension to the original shape from focus method. A strong texture is forced on imaged surfaces by the use of active illumination. The exact pattern of the projected illumination is determined through a careful Fourier analysis of all the optical effects involved in focus analysis. When the focus operator used is a 2-d Laplacian, the optimal illumination pattern is found to be a checkerboard whose pitch is the same size as the distance between adjacent elements in the discrete Laplacian kernel. This analysis also reveals the exact number of images required for accurate shape recovery. These results are experimentally verified using an optical microscope. Surfaces lacking in texture, such as, three-dimensional structures on silicon substrates and solder joints on circuit boards were used in the experiments. The results show that the derived illumination pattern is in fact optimal, facilitating accurate shape recovery of complex and pertinent industrial samples.

1 Introduction

Several passive methods have been proposed for recovering three-dimensional shapes of objects from their two-dimensional images, including, shape from shading, binocular stereo, structure from motion, and depth from focus/defocus. An important advantage of depth from focus is that, unlike stereo and motion, it is not confronted with the correspondence problem. As a result it is computational efficient, requiring only the use of simple focus measure operators. Several focus operators and depth estimation algorithms based on these operators have been suggested in the past [3] [4] [6] [9] [10] [11] [12] [13] [15] [17] [18] [19] [20]. An inherent weakness of focus-based methods, however, is that they require the imaged scene to have significant texture. Since defocusing is equivalent to low-pass filtering, depth from focus (or defocus) is effective only if the scene has high-frequency brightness variations. It is not a viable technique for smoothly shaded surfaces.

In [13], a shape from focus system was developed for microscopic objects. The high magnification of a microscope results in images that capture brightness variations caused by the micro-structure of the surface. Most surfaces that appear smooth and non-textured to the naked eye produce highly textured images under a microscope. Examples of such surfaces are paper, plastics, ceramics, etc. Microscopic shape from focus [13] was therefore demonstrated to be an effective approach, offering solutions to a variety to challenging shape inspection problems. However, there exist surfaces that are smooth at the micro-structure level and consequently do not produce sufficient texture even under a microscope. Examples are surfaces of solder joints and silicon wafers that also happen to be of great industrial import. For such samples, shape from focus has remained inadequate.

In [15], it was shown that textures can be forced on non-textured surfaces by using active illumination. An illumination pattern was projected on the scene and the resulting texture was used to estimate depth from defocus. In this case, the illumination pattern was selected in an arbitrary fashion. This does not guarantee the desired accuracy in computed depth. A critical problem therefore is determining the illumination pattern that would maximize the accuracy and robustness of shape from focus. In this paper, a solution to this problem is arrived at through a detailed Fourier analysis of the entire shape from focus system. First, theoretical models are developed for each optical phenomenon involved, including, the optical transfer function, defocusing, image sensing, and focus estimation. The derivation of the illumination pattern (or filter) is then posed an optimization problem in Fourier domain. The optimal illumination is one that maximizes sensitivity of the focus measure to depth variations, while minimizing the size of the focus operator to achieve high spatial resolution in computed depth. The analysis is carried further to determine the minimum number of images required for accurate depth estimation.

The resulting illumination filter was fabricated using micro-lithography and incorporated into the shape from focus system [13]. The illumination pattern is projected onto the sample via the same optics used to image the sample. This enables very precise registration of the illumination pattern with the sampling grid of the image sensor; light rays projected out through the imaging optics are subjected to the same geometric distortions as rays reflected back to the image sensor. Experiments have been conducted using samples that include textured and non-textured regions. Accurate and detailed depth maps of structures on silicon substrates and solder joints are reconstructed. The optimal illumination is shown to produce higher accuracy than other sub-optimal illuminations.

2 Shape from Focus

We begin with a brief overview of the microscope-based shape from focus system described in [13]. Fig.1 shows a schematic diagram of the various components of the system. The sample of interest (object) is placed on the translational stage of the microscope. The sample is magnified using the objective lens of the microscope and imaged by a standard CCD TV camera. In the original implementation [13], the sample is illuminated using the bright-field illumination of the microscope where light rays emitted by a source are projected onto the sample through the same objective lens used to image the sample. This results in uniform illumination over the entire field of view. The stage of the microscope is motorized and computer controlled. The sample is displaced in the z direction in increments of Δz and a set of images (typically 10 to 15) is taken. As the stage moves, each surface patch on the sample increases in image focus and then again decreases in focus. The Laplacian operator is applied to the image set to obtain a set of focus measures at each image point. The depth of a surface point is determined from its focus measure values by finding the peak of the focus measure function. The peak is localized by Gaussian interpolation of the three largest focus measures. This system was tested on a variety of complex samples, such as, via-hole fillings on ceramic substrates

*This research was conducted at the Center for Research in Intelligent Systems, Department of Computer Science, Columbia University. Minori Noguchi is a visiting researcher from the Production Engineering Research Laboratory, Hitachi Ltd., 292 Yoshida-cho, Totsuka-ku, Yokohama 244, Japan.

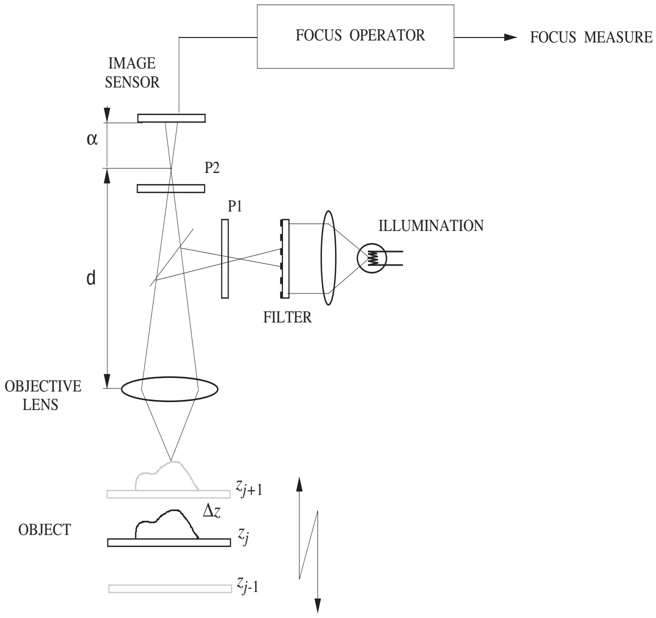


Figure 1: Microscopic shape from focus.

[13]. Such samples have substantial texture, enabling the recovery of very precise depth maps.

Here, we are concerned with surfaces that do not produce sufficient texture for focus analysis. To overcome this problem, we propose the projection of an illumination pattern on the imaged surface. This can be achieved by using a patterned filter in front of the microscope light source (see Fig.1). In addition, two polarization filters (p_1 and p_2), that are cross-polarized with respect to one another, are placed in front of the light source and the image sensor to minimize specularities in the sensed images. The projected illumination pattern has the effect of forcing a strong texture on the surface, and enables the original shape from focus method to once again produce accurate depth maps. The accuracy of computed depth clearly depends on the illumination pattern used and its relationship with other optical and computational components of the shape from focus system. Our goal is to model each component of the system, analyze the interplay between these components, and derive an illumination pattern (filter) that produces depth maps of the highest accuracy.

3 Modeling

There are five different elements, or components, of the shape from focus system that play a critical role in depth estimation. We briefly describe these components before we proceed to model them.

1. **Illumination Pattern:** The exact pattern used to illuminate the sample determines the final texture of the surface. The spatial and frequency characteristics of this texture determine the behavior of the focus measure function and hence the accuracy of depth estimation. It is the parameters of this component that we set out to optimize so as to achieve maximum depth accuracy.

2. **Optical Transfer Function:** The finite size of the lens aperture imposes restrictions on the range of spatial frequencies that are detectable by the imaging system. These restrictions play a critical role in the optimization of the illumination pattern.

3. **Defocusing:** The depth of a surface point is directly related to its defocus (or lack of it) on the image plane. It is this phenomenon that enables us to recover shape by focus analysis.

4. **Image Sensing:** The images used for shape recovery are of course discrete. The relationship between the continuous image formed on the sensor plane and the discrete image used in the computations is determined by the shape and spatial arrangement of sensing elements (pixels) on the image detector.

5. **Focus Operator:** The degree of focus/defocus in the image is estimated by using a focus operator. Such an operator is typically a high-pass filter and is applied to discrete images. Interestingly, the optimal illumination pattern is also dependent on the parameters of the focus operator used. In our case, a discrete Laplacian operator is implemented.

All the above components together determine the focus measure function, i.e. the relationship between focus measure values produced by the focus operator and the depth of the surface point. Therefore, the optimal illumination is viewed as one that maximizes the sensitivity and robustness of the focus measure function. To achieve this each component is modeled in spatial as well as Fourier domains.

3.1 Illumination Pattern

Before the parameters of the illumination pattern can be determined, an illumination model must be defined. Such a model must be generic in that it must subsume a large enough variety of possible illumination patterns. In defining the model, it is meaningful to take the characteristics of the other components into consideration. As we will describe shortly, the image sensor used has rectangular pixels arranged on a rectangular spatial grid. With this in mind, we define the following illumination model. The basic building block of the pattern is a rectangular illuminated patch, or cell, with uniform intensity:

$$i_c(x,y) = i_c(x,y;b_x,b_y) = {}^2\Pi(\frac{1}{b_x}x, \frac{1}{b_y}y) \qquad (1)$$

where, ${}^2\Pi()$ is the two-dimensional *Rectangular* function [2]. The unknown parameters of this illumination cell are b_x and b_y, the length and width of the cell.

This cell is assumed to be repeated on a two-dimensional grid defined as:

$$i_g(x,y) = i_g(x,y;t_x,t_y) = {}^2\mathrm{III}(\tfrac{1}{2}(\tfrac{1}{t_x}x + \tfrac{1}{t_y}y),\ \tfrac{1}{2}(\tfrac{1}{t_x}x - \tfrac{1}{t_y}y)) \qquad (2)$$

where, ${}^2\mathrm{III}()$ is the 2-dimensional *Shah* function [2], and t_x and t_y determine the periods of the grid in the x and y directions. Note that this grid is not rectangular but has vertical and horizontal symmetry on the $x-y$ plane. The final illumination pattern $i(x,y)$ is obtained by convolving the cell $i_c(x,y)$ with the grid $i_g(x,y)$:

$$i(x,y) = i(x,y;b_x,b_y,t_x,t_y) = i_c(x,y) * i_g(x,y) \qquad (3)$$

The exact pattern is therefore determined by four parameters, namely, b_x, b_y, t_x and t_y. In practice, the illumination pattern determined by the optimization is used to fabricate a filter with the same pattern. This filter is placed in front of the light source of the microscope (see Fig. 1). The above illumination grid is not as restrictive as it may appear upon initial inspection. For instance, the cell parameters b_x and b_y can each be stretched to obtain repeated illumination and non-illumination stripes in the horizontal and vertical directions, respectively. Alternatively, they can also be adjusted to obtain a checkerboard illumination pattern with large or small illuminated patches. The exact values for b_x, b_y, t_x and t_y will be evaluated by the optimization procedure described later.

The optimization procedure requires the analysis of each component of the shape from focus system in spatial domain as well as frequency domain (u, v). The Fourier transforms of the illumination cell, grid, and pattern are denoted as $I_c(u,v)$, $I_g(u,v)$, and $I(u,v)$, respectively, and found to be:

$$I_c(u,v) = I_c(u,v;b_x,b_y) = b_x \frac{\sin(\pi b_x u)}{\pi b_x u} \cdot b_y \frac{\sin(\pi b_y v)}{\pi b_y v} \tag{4}$$

$$I_g(u,v) = I_g(u,v;t_x,t_y) = {}^2\mathrm{III}(\tfrac{1}{2}(t_x u + t_y v),\ \tfrac{1}{2}(t_x u - t_y v)) \tag{5}$$

$$I(u,v) = I(u,v;b_x,b_y,t_x,t_y) = I_c(u,v) \cdot I_g(u,v) \tag{6}$$

3.2 Optical Transfer Function

Adjacent points on the illuminated surface reflect light waves that interfere with each other to produce diffraction effects. As a result, the frequency of the reflected wave increases with the angle of reflection. The exact frequency radiated in any particular direction depends on the spatial frequency of the surface texture, the wavelength of incident light, and the angle of reflection. Since the lens aperture of the imaging system is of finite radius, it does not capture the higher frequencies radiated by the surface. This effect places a limit on the optical resolution of the imaging system, which is characterized by the optical transfer function (see [1] for details):

$$\begin{aligned} O(u,v) &= O(u,v;a,d') \qquad (7) \\ &= \begin{cases} (\frac{a}{\lambda d'})^2(\gamma - \sin\gamma)\ , & \sqrt{u^2+v^2} \le \frac{2a}{\lambda d'} \\ 0\ , & \sqrt{u^2+v^2} > \frac{2a}{\lambda d'} \end{cases} \\ &\text{where } \gamma = \cos^{-1}(\tfrac{\lambda d'}{a}\tfrac{\sqrt{u^2+v^2}}{2})\ . \end{aligned}$$

where, (u,v) is the spatial frequency of the two-dimensional surface texture, a is the radius of the lens aperture, d' is the distance of the surface from the lens, and λ is the wavelength of incident light. It is clear from the above expression that only spatial frequencies below the limit $\frac{2a}{\lambda d'}$ will be imaged by the optical system.

3.3 Defocusing

The defocus function is described in detail in previous work (see [1] [7], for example). Let α be the distance between the focused image of a surface point and its defocused image formed on the sensor plane. The light energy radiated by the surface point and collected by the imaging optics is uniformly distributed over a circular patch on the sensor plane. This patch, also called the *pillbox*, is the defocusing function:

$$h(x,y) = h(x,y;\alpha,a,d) = \frac{K}{\pi a^2}\Pi(\frac{d}{2a\alpha}\sqrt{x^2+y^2}) \tag{8}$$

where, once again, a is the radius of the lens aperture. Parameter d is the distance between the principal point of the lens and the focused image. K represents the radiance, or brightness, of the surface point.

In Fourier domain, the above defocus function is given by:

$$H(u,v) = H(u,v;\alpha,a,d) = \frac{\alpha K}{\pi a d\sqrt{u^2+v^2}} J_1(\frac{2\pi a\alpha}{d}\sqrt{u^2+v^2}) \tag{9}$$

where J_1 is the first-order Bessel function [1].

3.4 Image Sensing

We assume the image sensor to be a typical CCD TV camera. Such a sensor can be modeled as a rectangular array of rectangular sensing elements (pixels). The quantum efficiency [8] of each pixel is assumed to be uniform over the area of the pixel. Let $m(x,y)$ be the continuous image formed on the sensor plane. The finite pixel area has the effect of averaging the continuous image $m(x,y)$. In spatial domain, the averaging function is the rectangular cell:

$$s_c(x,y) = s_c(x,y;w_x,w_y) = {}^2\,\Pi(\frac{1}{w_x}x, \frac{1}{w_y}y) \tag{10}$$

where, w_x and w_y are the length and width of the pixel, respectively. The discrete image is obtained by sampling the image obtained by convolving $m(x,y)$ with $s_c(x,y)$. This sampling function is a rectangular grid:

$$s_g(x,y) = s_g(x,y;p_x,p_y,\varphi_x,\varphi_y) = {}^2\mathrm{III}(\tfrac{1}{p_x}(x-\varphi_x), \tfrac{1}{p_y}(y-\varphi_y)) \tag{11}$$

where, p_x and p_y are spacings between discrete samples in the two spatial dimensions, and (φ_x, φ_y) is phase shift of the grid. The final discrete image is therefore:

$$m_d(x,y) = (s_c(x,y) * m(x,y)) \cdot s_g(x,y) \tag{12}$$

The parameters w_x, w_y, p_x, and p_y are all determined by the particular image sensor used. These parameters are therefore known and their values are substituted after the optimization is done. On the other hand, the phase shift (φ_x, φ_y) of the sampling function is with respect to the illumination pattern and will be determined by the optimization process.

In Fourier domain, the above averaging and sampling functions are:

$$S_c(u,v) = S_c(u,v;w_x,w_y) = w_x\frac{\sin(\pi w_x u)}{\pi w_x u} \cdot w_y \frac{\sin(\pi w_y v)}{\pi w_y v} \tag{13}$$

$$\begin{aligned} S_g(u,v) &= S_g(u,v;p_x,p_y,\varphi_x,\varphi_y) \qquad (14) \\ &= {}^2\mathrm{III}(p_x u, p_y v) \cdot e^{-i2\pi(\varphi_x u + \varphi_y v)} \end{aligned}$$

The final discrete image is:

$$M_d(u,v) = (S_c(u,v) \cdot M(u,v)) * S_g(u,v) \tag{15}$$

3.5 Focus Operator: Laplacian

Since defocusing has the effect of suppressing high-frequency components in the focused image, it is desirable that the focus operator respond to high frequencies in the image. A suitable focus operator is the Laplacian[1]. In spatial domain, the discrete Laplacian is:

$$\begin{aligned} l(x,y) &= l(x,y;q_x,q_y) \qquad (16) \\ &= 4\delta(x)\cdot\delta(y) - [\delta(x)\cdot\delta(y-q_y) + \delta(x)\cdot\delta(y+q_y) \\ &\quad +\delta(x-q_x)\cdot\delta(y) + \delta(x+q_x)\cdot\delta(y)] \\ &= 2[\delta(x) - \mathrm{II}(\tfrac{1}{2q_x}x)]\cdot\delta(y) + 2[\delta(y) - \mathrm{II}(\tfrac{1}{2q_y}y)]\cdot\delta(x) \end{aligned}$$

Here, q_x and q_y are the spacings between neighboring elements of the discrete Laplacian kernel. In the optimization, these spacings will be related to the illumination parameters. The Fourier transform of the discrete Laplacian is:

$$\begin{aligned} L(u,v) &= L(u,v;q_x,q_y) \qquad (17) \\ &= 2(1-\cos(2\pi q_x u)) * \delta(u) + 2(1-\cos(2\pi q_y v)) * \delta(v) \\ &= 4 - 2\cos(2\pi q_x u) - 2\cos(2\pi q_y v) \end{aligned}$$

3.6 Focus Measure Function

The focus measure function relates the output of the Laplacian operator to the defocus parameter α. Its analytical form therefore depends on all the five components modeled above. Note that the illumination pattern $(i_c * i_g)$ is projected through the same optics that is used for image formation. Consequently, this pattern is also subject to the limits imposed by the optical transfer function o and

[1] In [13], a modified version of the Laplacian operator was used to overcome the problem of zero-crossings caused by the unknown surface texture. Here, the texture is determined by the illumination pattern and hence the Laplacian is adequate.

the defocus function h. Therefore, the texture projected on the physical surface of the sample is:

$$i(x,y;b_x,b_y,t_x,t_y) * o(x,y;a,d') * h(x,y;\alpha,a,d) \tag{18}$$

This texture is again reflected by the viewed surface and projected by the optics back onto the image plane to produce the discrete image:

$$\begin{aligned}\{i(x,y;b_x,b_y,t_x,t_y) * o(x,y;a,d')^{*2} * h(x,y;\alpha,a,d)^{*2} \\ *s_c(x,y;w_x,w_y)\} \cdot s_g(x,y;p_x,p_y,\varphi_x,\varphi_y)\end{aligned} \tag{19}$$

where, $f^{*2} = f * f$. The final focus measure function $g(x,y)$ is the result of applying the discrete Laplacian to the above discrete image:

$$\begin{aligned} g(x,y) &= \{(i(x,y;b_x,b_y,t_x,t_y) * o(x,y;a,d')^{*2} \\ &\quad *h(x,y;\alpha,a,d)^{*2} * s_c(x,y;w_x,w_y)) \\ &\quad \cdot s_g(x,y;p_x,p_y,\varphi_x,\varphi_y)\} * l(x,y;q_x,q_y) \\ &= \{(i * o^{*2} * h^{*2} * s_c) \cdot s_g\} * l \end{aligned} \tag{20}$$

Since the distance between adjacent weights of the Laplacian kernel must be integer multiples of the period of the image sampling function s_g, the above expression can be rearranged as:

$$\begin{aligned} g(x,y) &= (i * o^{*2} * h^{*2} * s_c * l) \cdot s_g \\ &= g_0 \cdot s_g \end{aligned} \tag{21}$$

where, $g_0 = i * o^{*2} * h^{*2} * s_c * l$. The same can be expressed in Fourier domain as:

$$\begin{aligned} G(u,v) &= (I \cdot O^2 \cdot H^2 \cdot S_c \cdot L) * S_g \\ &= G_0 * S_g \end{aligned} \tag{22}$$

The above expression will be used in the following section to determine the optimal illumination pattern.

4 Optimization

We will assume in the following discussion that the projected illumination is the primary cause for surface texture and is stronger than the natural texture of the surface. This enables us to apply our results not only to non-textured surfaces but also textured ones. The illumination optimization problem is formulated as follows: Establish closed-form relationships between the illumination parameters (b_x, b_y, t_x, t_y), sensor parameters $(w_x, w_y, p_x, p_y, \varphi_x, \varphi_y)$, and discrete Laplacian parameters (q_x, q_y) so as to maximize the sensitivity, robustness, and spatial resolution of the focus measure function $g(x,y)$. High sensitivity implies that a small variation in the degree of focus results in a large variation in $g(x,y)$. This would ensure high depth estimation accuracy in the presence of image noise, i.e. high signal-to-noise ratio. By robustness we mean that all pixels with the same degree of defocus produce the same focus measure value independent of their location on the image plane. Lastly, high spatial resolution is achieved by minimizing the size of the Laplacian focus operator. This ensures that rapid depth variations (surface discontinuities) can be detected with high accuracy.

For lack of space, we shall avoid mathematical details and confine our discussion to the main results. In order to minimize smoothing effects and maximize spatial resolution of computed depth, the support (or span) of the discrete Laplacian must be as small as possible. This in turn requires that the frequency of the illumination pattern be as high as possible. However, the optical transfer function described in Section 3.2 imposes limits on the highest frequency that can be imaged by the optical system. This maximum allowable frequency is $\frac{2a}{\lambda d'}$. With this in mind, let us examine the Fourier transform of the illumination pattern. Since the pattern is periodic, its Fourier transform must be discrete. It may have a zero-frequency component, but this can be safely ignored since the Laplacian operator, being a second-order derivative operator, will eventually remove any zero-frequency component in the final image. Our objective then is to maximize the fundamental spatial frequencies $(1/t_x, 1/t_y)$ of the illumination pattern. In order to maximize illumination frequency while maintaining high detectability, we must have $\sqrt{(1/t_x)^2 + (1/t_y)^2}$ close to the optical limit $\frac{2a}{\lambda d'}$. This in turn pushes all higher frequencies in the illumination pattern outside the optical limit. What we are left with is a surface texture whose image has only the fundamental frequencies $(\pm 1/t_x, \pm 1/t_y)$. As a result, these are the only frequencies we need consider in our analysis of the focus measure function $G(u,v)$.

Before we consider the final measure $G(u,v)$, we examine $G_0(u,v)$ the focus measure prior to image sampling. For the reasons given above, $G_0(u,v)$ is reduced to four discrete spikes at $(1/t_x, 1/t_y)$, $(1/t_x, -1/t_y)$, $(-1/t_x, 1/t_y)$ and $(-1/t_x, -1/t_y)$. Since all components (I, O, H, S_c and L) of G_0 are reflection symmetric about $u = 0$ and $v = 0$, we have:

$$G_0(\frac{1}{t_x}, \frac{1}{t_y}) = G_0(\frac{1}{t_x}, -\frac{1}{t_y}) = G_0(-\frac{1}{t_x}, \frac{1}{t_y}) = G_0(-\frac{1}{t_x}, -\frac{1}{t_y}) \tag{23}$$

and:

$$\begin{aligned} G_0(\frac{1}{t_x}, \frac{1}{t_y}) &= I(\frac{1}{t_x}, \frac{1}{t_y}; b_x, b_y, t_x, t_y) \\ &\quad \cdot O^2(\frac{1}{t_x}, \frac{1}{t_y}; a, d) \cdot H^2(\frac{1}{t_x}, \frac{1}{t_y}; \alpha, a, d) \\ &\quad \cdot S_c(\frac{1}{t_x}, \frac{1}{t_y}; w_x, w_y) \cdot L(\frac{1}{t_x}, \frac{1}{t_y}; q_x, q_y) \cdot \end{aligned} \tag{24}$$

Therefore, in Fourier domain the focus measure function prior to image sampling reduces to:

$$\begin{aligned} G_0(u,v) &= G_0(\frac{1}{t_x}, \frac{1}{t_y}) \\ &\quad \cdot \{\delta(u - \frac{1}{t_x}) \cdot \delta(v - \frac{1}{t_y}) + \delta(u + \frac{1}{t_x}) \cdot \delta(v - \frac{1}{t_y}) \\ &\quad + \delta(u - \frac{1}{t_x}) \cdot \delta(v + \frac{1}{t_y}) + \delta(u + \frac{1}{t_x}) \cdot \delta(v + \frac{1}{t_y})\} \end{aligned} \tag{25}$$

The function $g_0(x,y)$ in image (spatial) domain, is simply the inverse Fourier transform of $G_0(u,v)$:

$$g_0(x,y) = G_0(\frac{1}{t_x}, \frac{1}{t_y}) \cdot \{4 \cos 2\pi \frac{1}{t_x} x \cdot \cos 2\pi \frac{1}{t_y} y\} \tag{26}$$

It is seen that $g_0(x,y)$ is the product of cosine functions weighted by the coefficient $G_0(1/t_x, 1/t_y)$. The defocus function h has the effect of reducing the coefficient $G_0(1/t_x, 1/t_y)$ in the focus measure function $g_0(x,y)$. Clearly, the sensitivity of the focus measure to depth (or defocus) is optimized by maximizing the coefficient $G_0(1/t_x, 1/t_y)$ with respect to the unknown parameters of the system. This optimization procedure can be summarized as:

$$\frac{\partial}{\partial t_x} G_0(\frac{1}{t_x}, \frac{1}{t_y}) = 0, \quad \frac{\partial}{\partial t_y} G_0(\frac{1}{t_x}, \frac{1}{t_y}) = 0, \tag{27}$$

$$\frac{\partial}{\partial b_x} G_0(\frac{1}{t_x}, \frac{1}{t_y}) = 0, \quad \frac{\partial}{\partial b_y} G_0(\frac{1}{t_x}, \frac{1}{t_y}) = 0, \tag{28}$$

$$\frac{\partial}{\partial q_x} G_0(\frac{1}{t_x}, \frac{1}{t_y}) = 0, \quad \frac{\partial}{\partial q_y} G_0(\frac{1}{t_x}, \frac{1}{t_y}) = 0, \tag{29}$$

Since t_x and t_y show up in all the components in (24), the first two partial derivatives (equation (27)) are difficult to evaluate. Fortunately, the derivatives in (28) and (29) are sufficient to obtain the following relationships between the system parameters:

$$b_x = \frac{1}{2} t_x, \quad b_y = \frac{1}{2} t_y \tag{30}$$

$$q_x = \frac{1}{2} t_x, \quad q_y = \frac{1}{2} t_y \tag{31}$$

If the above relations are satisfied, the sensitivity of $g(x,y)$ to defocus is maximized. Next, we examine the spatial robustness of $g(x,y)$. Imagine the imaged surface to be planar and parallel to the image sensor. In this case, we would like our image sampling to produce the same absolute value of $g(x,y)$ at all discrete sampling points on the image. This entails relating the illumination and sensing parameters so as to facilitate careful sampling of the product of cosines in (26). Note that the final focus measure function is:

$$g(x,y) = g_0 \cdot s_g = G_0(\frac{1}{t_x}, \frac{1}{t_y}) \cdot \{4\cos 2\pi \frac{1}{t_x} x \cdot \cos 2\pi \frac{1}{t_y} y\} \cdot {}^2\mathrm{III}(\frac{1}{p_x}(x-\varphi_x), \frac{1}{p_y}(y-\varphi_y)) \tag{32}$$

All samples of $g(x,y)$ have the same *absolute* value when the two cosines in the above expression are sampled at their peak values. Such a sampling is possible when:

$$p_x = \frac{1}{2} t_x, \quad p_y = \frac{1}{2} t_y \tag{33}$$

and:

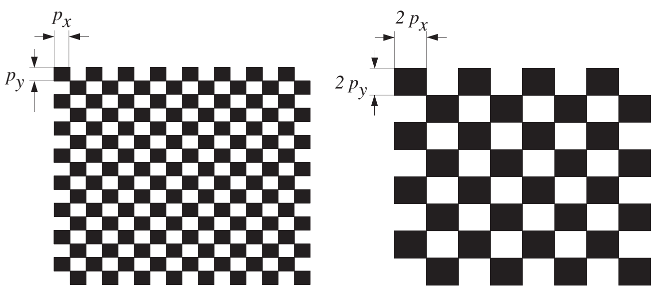
$$\varphi_x = 0, \quad \varphi_y = 0 \tag{34}$$

Alternatively, the cosines can be sampled with a period of $\pi/2$ and phase shift of $\pi/4$. This yields the second solution:

$$p_x = \frac{1}{4} t_x, \quad p_y = \frac{1}{4} t_y, \tag{35}$$

$$\varphi_x = \pm\frac{1}{8} t_x, \quad \varphi_y = \pm\frac{1}{8} t_y. \tag{36}$$

The solutions derived above are used to fabricate illumination filters that are placed in the path of the microscope illumination. This generates the desired texture on the sample surfaces. As we will see, the resulting images can be used to recover precise shape information. The above equations give two solutions, both are checkerboard illumination patterns but differ in their fundamental frequencies, size of the illumination cell, and phase shift with respect to the image sensor. Equations (30), (31), (33), and (34) yield the filter pattern shown in Fig.2(a). In this case, the filter and sensor are registered with zero phase shift, and the illumination cell has the same size and shape as the sensor pixels. The second filter, shown in Fig.2(b), is obtained using the sampling solutions (35) and (36), yielding a filter pattern with illumination cell two times the size of the sensor pixel and phase shift of half the pixel size.



(a) (b)

Figure 2: Optimal illumination filter patterns: (a) $t_x = 2p_x, t_y = 2p_y, \varphi_x = 0, \varphi_y = 0$; and (b) $t_x = 4p_x, t_y = 4p_y, \varphi_x = 1/8t_x, \varphi_y = 1/8t_y$. Here, (t_x, t_y) is the illumination period, (p_x, p_y) is the pixel size, and (φ_x, φ_y) is the illumination phase shift with respect to the image sensing grid.

5 Implementation

The optimal illumination filter shown in Fig. 2(b) was incorporated into the shape from focus system developed in [13]. In addition, cross-polarized filters were fixed in front of the illumination filter and the image sensor to filter out strong specular reflections from the sample. A set of sample images are obtained by automatically moving the microscope stage in increments of $\Delta z = z_i - z_{i-1}$. The Laplacian focus operator is applied to each image to obtain a set of focus measure values at each image point (x,y); the number of focus measures equals the number of images taken. The discrete stage position z_j that yields the maximum focus measure value at an image point, can be used as an approximation of the depth of the corresponding surface point. A more accurate depth estimate z is obtained by applying Gaussian interpolation [13] to the three focus measures corresponding to z_{j-1}, z_j, and z_{j+1}.

It was shown in [13] that at least 3 focus measures are needed for depth estimation by Gaussian interpolation. In the active illumination system proposed here, the fundamental frequency of the surface texture is determined by the illumination filter used and is simply $\sqrt{(\frac{1}{t_x})^2 + (\frac{1}{t_y})^2}$. For this frequency the defocus measure has a zero-crossing at defocus value α' such that $\frac{2\pi a\alpha'}{d}\sqrt{(\frac{1}{t_x})^2 + (\frac{1}{t_y})^2} = 3.83$. This gives us an upper limit on the usable defocus range for any point on the surface, which is, $-\alpha' \leq \alpha \leq \alpha'$. Therefore, on the image side of the optics, we need to obtain (for any surface point) at least 3 images within the above defocus range. This gives us the following maximum distance between consecutive focused images on the image side of the optics:

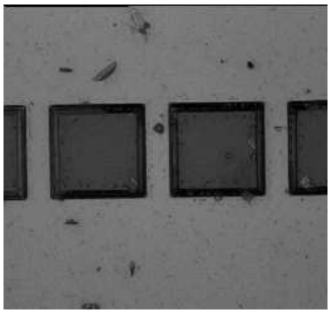
$$\Delta z' \leq \frac{2\alpha'}{4} = \frac{1}{2} \cdot 3.83 \cdot \frac{d}{2\pi a \sqrt{(\frac{1}{t_x})^2 + (\frac{1}{t_y})^2}} \tag{37}$$

This distance on the image side is related to the maximum allowable incremental displacement (between consecutive images) on the sample side of the optics by the magnification M of the objective lens used:

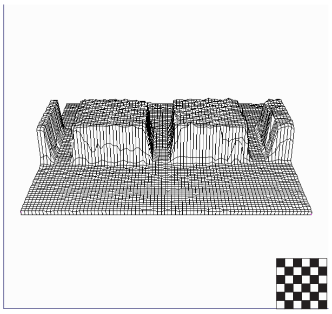
$$\Delta z \simeq \Delta z' \frac{1}{M^2} \tag{38}$$

In our experiments, the magnification of the objective lens is $M = 20$, the ratio $a/d = 0.025$, and illumination parameters are $t_x = 44\ \mu m$, $t_y = 52\ \mu m$. Using these values in eqs.(37) and (38) we get the maximum allowable stage displacement $\Delta z \leq 0.5\ \mu m$. Experiments reported in [14] illustrate that $\Delta z = 0.5\ \mu m$ does in fact produce the best results. Further decreasing Δz does not significantly improve the accuracy of the depth maps.

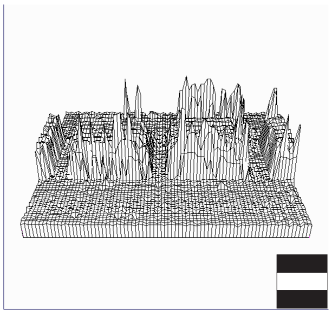
Figs. 3 and 4 illustrate the advantage of using the optimal illumination over sub-optimal and uniform illumination. The sample in Fig. 3 has rectangular structures fabricated on a smooth silicon wafer. Silicon wafer inspection is of great relevance in a variety of chip manufacturing processes. The surface of the wafer is very smooth, resulting in images (see Fig. 3(a)) that are more or less textureless. This renders the original shape from focus system [13] ineffective (see Fig. 3(d)). A total of 16 images were taken for this sample. The optimal illumination (Fig. 3(b)) produces very accurate shape information that is superior to that produced by sub-optimal filters (Fig. 3(c)). Similar results are obtained for the solder joint sample shown in Fig. 4. For this sample, an objective lens with $M = 10$ was used and a total of 23 sample images were taken. This sample exhibits noticeable texture under a microscope. However, the texture is not consistent over the entire surface. As a result, active illumination once again produces the best results. Solder shape inspection has remained a challenging and unresolved industrial problem. These results indicate that the proposed technique may provide an effective solution to this important problem. In all of the above experiments, the cross-polarized filters were effective in cutting out specular reflections, avoiding image saturation and enabling the computation of robust focus measure values.



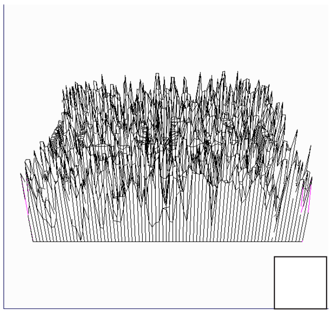
(a) Camera image



(b) With optimal illumination

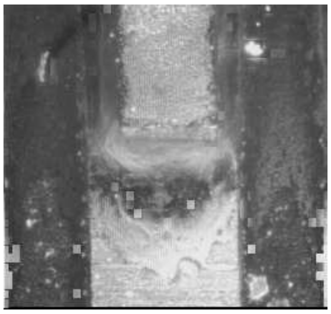


(c) With sub-optimal illumination

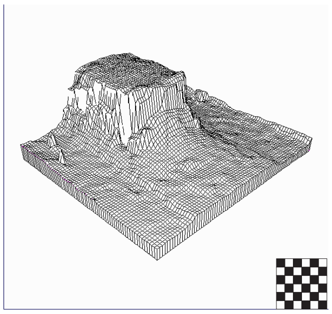


(d) No active illumination

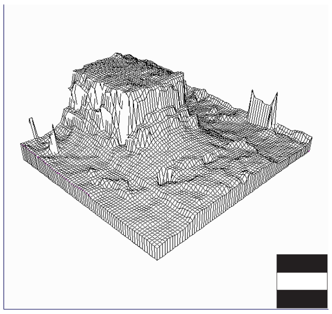
Figure 3: Depth maps of rectangular structures on a smooth silicon substrate. The structures are approximately $13\mu m$ tall. The insets illustrate the illumination filters used.



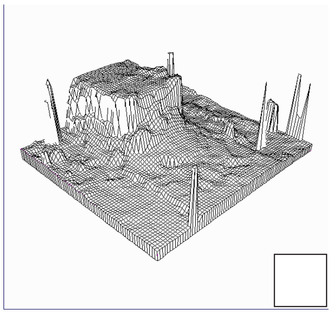
(a) Camera image



(b) With optimal illumination



(c) With sub-optimal illumination



(d) No active illumination

Figure 4: Depth maps of a solder joint on a circuit board. The solder joint is approximately $150\mu m$ high and $100\mu m$ wide.

6 Conclusion

In this paper, the shape from focus method has been enhanced by the use of active illumination. The optimal illumination pattern was determined by modeling five major components of the shape from focus system: (a) optical transfer function; (b) defocus function; (c) image sensing; (d) focus operator; and (e) the illumination pattern. The parameters of the illumination pattern were analytically derived to maximize the accuracy and spatial resolution of computed depth. Two checkerboard illumination patterns were found to satisfy the desired constraints. Of these the pattern with larger illumination period was implemented using an optical filter that is fabricated by micro-lithography. The fundamental frequency of the illumination pattern, or the surface texture, was also used to determine the minimum number of images needed for accurate shape recovery. Experiments were conducted on samples of silicon substrates and solder joints. These are examples of relevant industrial samples that have smooth textureless surfaces. The optimal illumination filter enabled the extraction of accurate shape information. These results were compared with those obtained using sub-optimal and uniform illuminations.

References

[1] M. Born and E. Wolf, *Principles of Optics,* London:Pergamon, 1965.

[2] R. N. Bracewell, *The Fourier Transform and Its Applications,* McGraw-Hill, 1965.

[3] T. Darrell and K. Wohn, *Pyramid Based Depth from Focus,* Proc. CVPR, pp. 504-509, 1988.

[4] J. Ens and P. Lawrence, *A Matrix Based Method For Determining Depth From Focus,* Proc. CVPR, pp. 600-606, 1991.

[5] J. Goodman, *Introduction to Fourier Optics,* McGraw-Hill, San Fransisco, 1968.

[6] P. Grossmann, *Depth from Focus,* Pattern Recognition Letters, Vol. 5, pp. 63-69, 1987.

[7] B.K.P. Horn, *Focusing,* MIT Artificial Intelligence Laboratory, Memo No. 160, May, 1968.

[8] B. K. P. Horn, *Robot Vision,* MIT Press, 1986.

[9] R.A. Jarvis, *Focus optimization criteria for computer image processing,* Microscope, Vol. 24, No. 2, pp. 163-180, 1976.

[10] A. Krishnan and N. Ahuja, *Range Estimation From Focus Using a Non-frontal Imaging Camera,* Proc. AAAI Conference, pp.830-835, July, 1993.

[11] E. Krotkov, *Focusing,* International Journal of Computer Vision, Vol. 1, pp. 223-237, 1987.

[12] G. Ligthart and F. Groen, *A comparison of different autofocus algorithms,* Proc. of International Conference on Pattern Recognition, pp. 597-600, 1982.

[13] S. K. Nayar and Y. Nakagawa, *Shape from Focus: An Effective Approach for Rough Surfaces,* , Proc. IEEE Intl. Conf. on Robotics and Automation, pp. 218-225, 1990.

[14] M. Noguchi and S. K. Nayar, *Recovering Microscopic Shapes Using Focus Analysis and Active Illumination,* Technical Report, CUCS-024-94, February, 1994.

[15] A. Pentland, *A New Sense for Depth of Field,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, No. 4, pp. 523-531, July 1987.

[16] A. Pentland, *Simple Range Cameras Based on Focal Error,* Proc. Intl. Simp. on Advanced Image-acquisition Technology, November 1991.

[17] M. Subbarao, *Efficient Depth Recovery through Inverse Optics,* Machine Vision for Inspection and Measurement, edited by H. Freeman, Academic Press, 1989.

[18] M. Subbarao and G. Surya, *Depth from Defocus: A Spatial Domain Approach,* Technical Report 92.120.45, SUNY, Stony Brook, December 92.

[19] J.M. Tenenbaum, *Accommodation in Computer Vision,* Ph.D. Thesis, Stanford University, 1970.

[20] Y. Xiong and S. A. Shafer, *Depth from Focusing and Defocusing,* Proc. CVPR, pp. 68-73, 1993.